Length Frequency

Alex J. Benecke

December 19, 2017

Data Preparation

I will now compare the length frequency distribution for largemouth bass obtained in the nearshore electrofishing survey during 2013 - 2016.

```
lmb <- read.csv("Data/Clean-Data/2012-2017_nearshore-survey-largemouth-bass_CLEAN.csv") %>%
    arrange(Year, FID, Length)
lmb$fyr <- as.factor(lmb$fyr)
```

```
unique(lmb$Year) ### See that there is no 2013
```

```
## [1] 2012 2013 2014 2015 2016 2017
```

Lets create a new variable for 20 mm length bins.

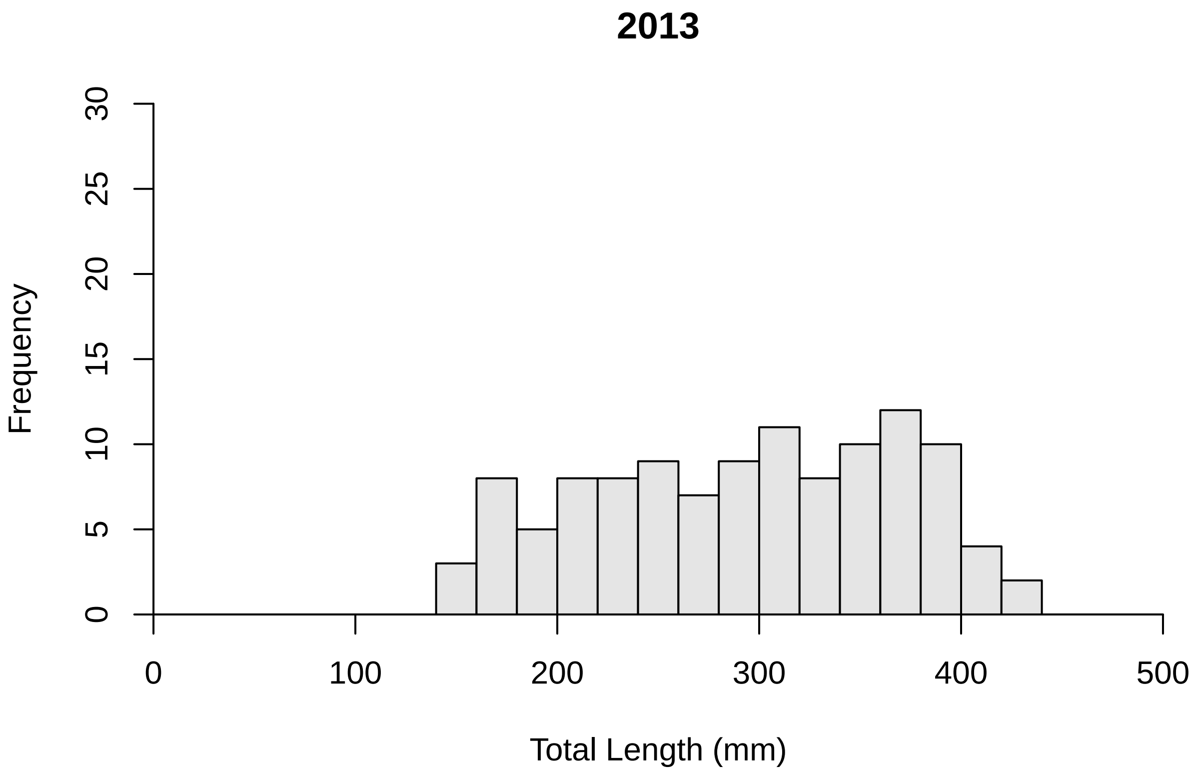
```
lmb %<>% mutate(lcat20 = lencat(Length, w = 20))
```

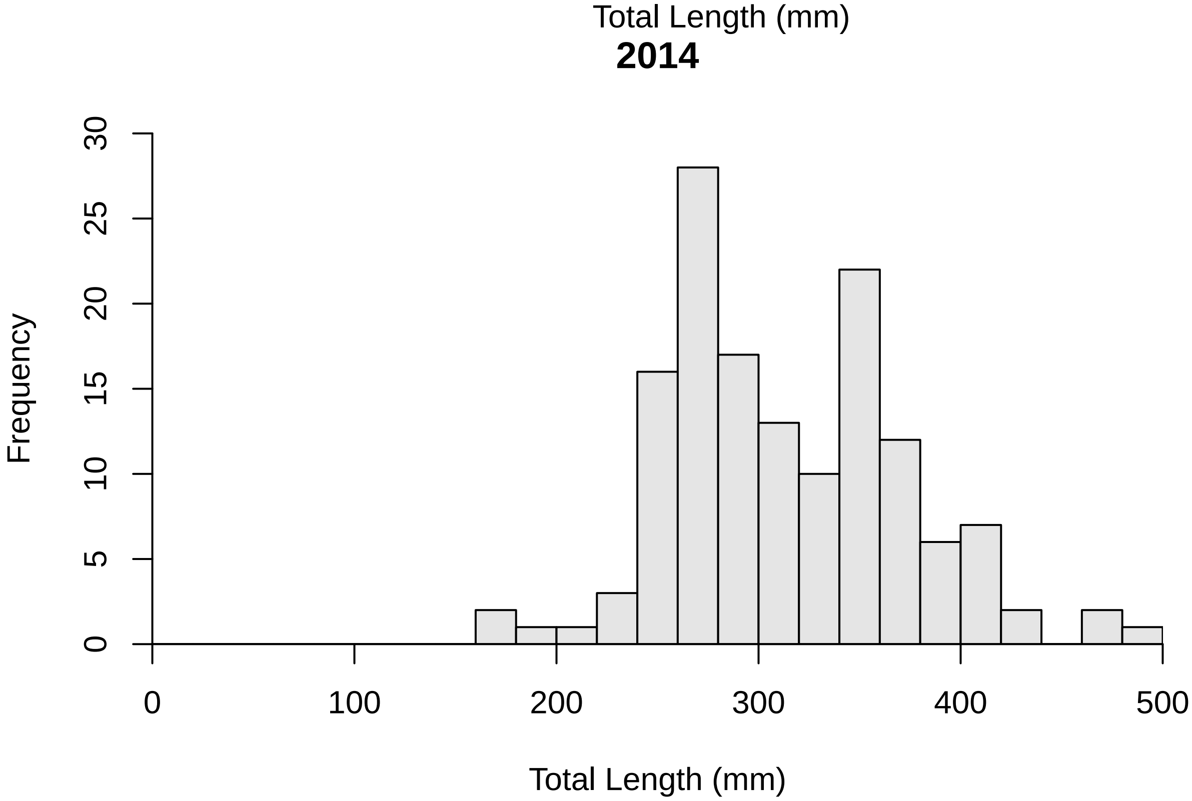
Now I want to separate out the years. I will throw out the year 2012 because samples from this years were not collected using the same procedures as in subsequent years. Only large LMB from 2012 had length weigh data.

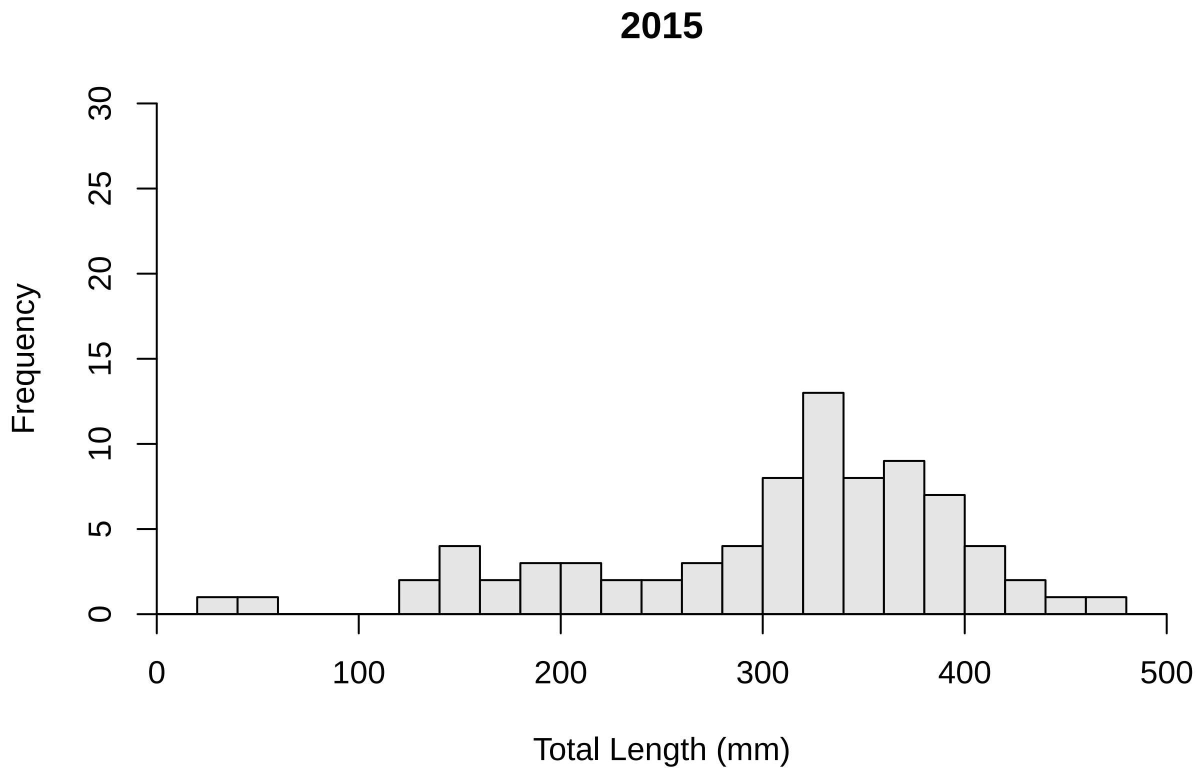
```
lmb.12 <- filter(lmb, Year == 2012)
# 1-8-2018#write.csv(lmb.12, file = 'Data/Clean-Data/minor-data/lmb.12.csv')
lmb.13 <- filter(lmb, Year == 2013)
# 1-8-2018#write.csv(lmb.13, file = 'Data/Clean-Data/minor-data/lmb.13.csv')
lmb.14 <- filter(lmb, Year == 2014)
# 1-8-2018#write.csv(lmb.14, file = 'Data/Clean-Data/minor-data/lmb.14.csv')
lmb.15 <- filter(lmb, Year == 2015)
# 1-8-2018#write.csv(lmb.15, file = 'Data/Clean-Data/minor-data/lmb.15.csv')
lmb.16 <- filter(lmb, Year == 2016)
# 1-8-2018#write.csv(lmb.16, file = 'Data/Clean-Data/minor-data/lmb.16.csv')
lmb.17 <- filter(lmb, Year == 2017)
# 1-10-2018#write.csv(lmb.17, file =
# 'Data/Clean-Data/minor-data/lmb.17.csv')
```

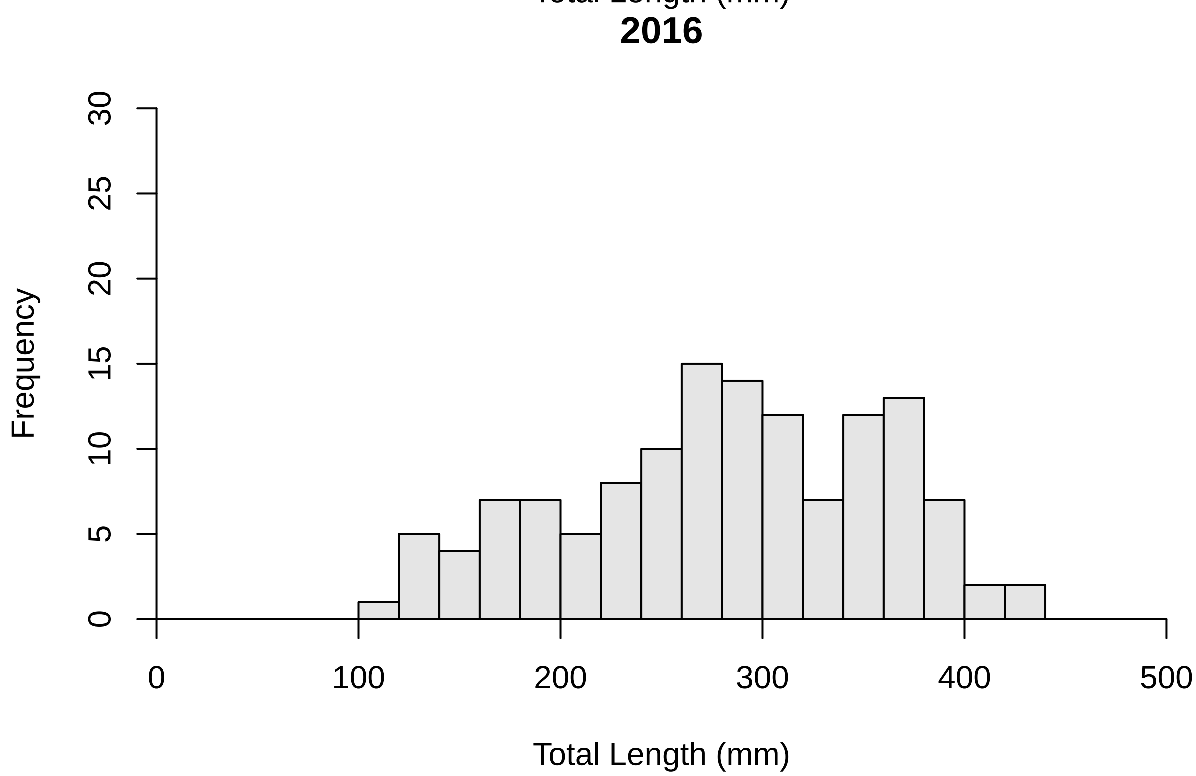
Length Frequency Distribution

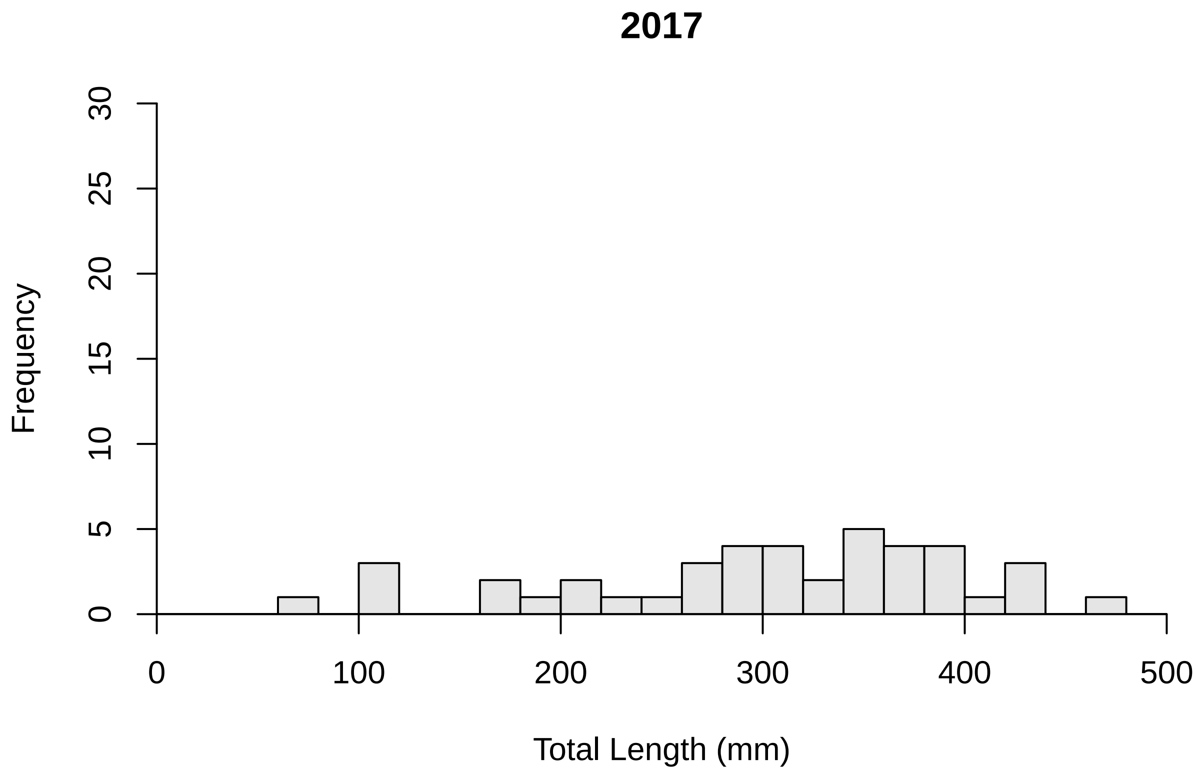
Lets view a quick histogram of the frequency of fish in each length bin.






2015
Frequency
0
5
10
15
20
25
30
0
100
200
300
400
500
Total Length (mm)


2016
Frequency
0
5
10
15
20
25
30
0
100
200
300
400
500
Total Length (mm)



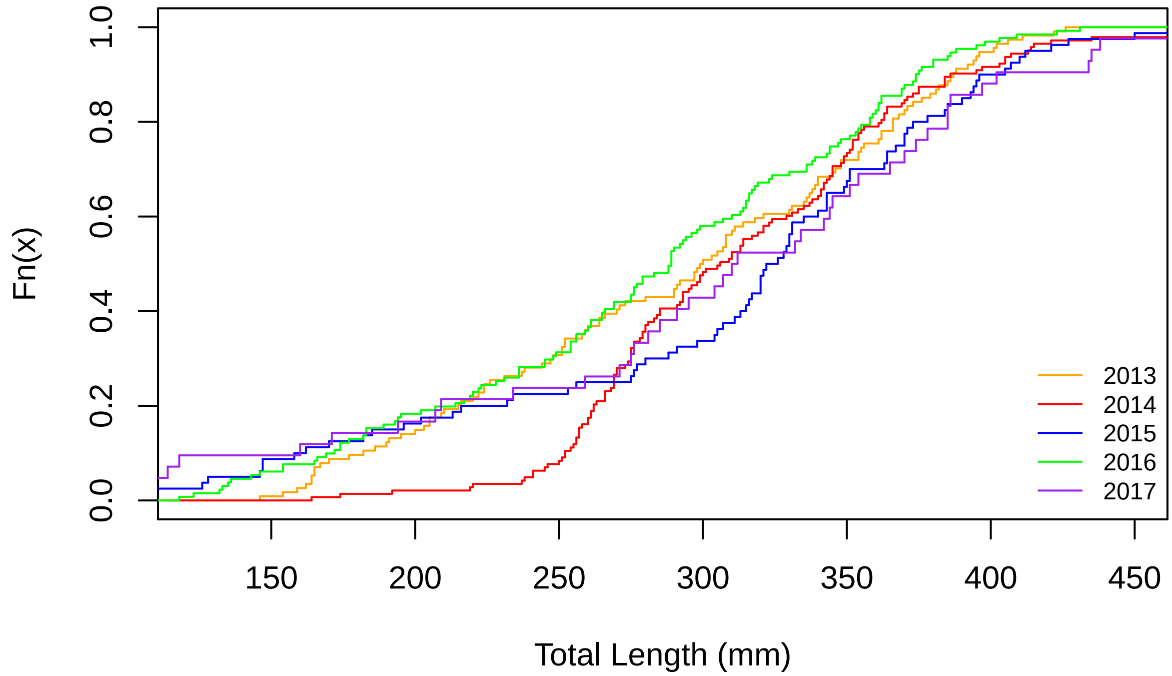
I can tell that small fish (<100-250mm) are definitely not being sampled consistently. 2014 and 2016 look more or less as I would expect with a stable largemouth bass population (2014 more so than 2016) with a peak and more or less gradual decline in the number of fish with increasing length. The peak for 2015 appears to be shifted right (320mm) more so than for 2014 and 2016 (260mm). The year 2013 Appears to be more or less flat but it's highest peak is at (360mm) before declining. As for 2017 I think this represents inadequate sampling I see a peak with much smaller fish (100mm) than other years and some length bins with only 1 fish. the peak for 2017 is around 340mm.

I would have to say from these graphs that the length frequency distribution of largemouth bass is unstable. However, I am tempted to think that there may be some aspect of sampling bias playing a role in this. 2017 was not sampled fully or by the same gear or crew (University of Toledo and old shock boat) due to staff shortages so I think I may throw this year out (missing 4 annual sites but has 2 rotating sites). 2013 is also missing several sites (6, 12, 19) although one site is NA so this could be one of those three. However, 2013 does seem to have a good sample size (n = 114) also site 19 was only sampled in 2017 and site 6 never seemed to provide alot of largemouth bass. So I will keep the year 2013 in at least for now. 2015 was also incompletely sampled missing two annual sotes (10 and 19) but with one rotating site for a total of nine sites. 2015 has a lower sample size than other years (n=80) except for 2017. However, since I dont't have any years in which all sites were sampled and the same number of sites were sampled in 2015 as in 2014 and 2016 I cannot throw out this year. Although, just looking at the data from 2015 it seems for whatever reason we have a significantly smaller proportion of smaller (stock and substock) fish.

length frequency all years 2013 - 2017

Cumulative Frequencies

Lets look at the empirical cumulative distribution function (ECDF). This is the proportion of fish less than each observed length. This should help me compare the length frequency distributions between years.



Compare Length Frequency Between Years

Kolmogorov-Smirnov Test

```
##  [1] 0.23721016 0.19517544 0.09997322 0.14160401 0.19003497 0.24747771
##  [7] 0.20313020 0.24265267 0.13630952 0.17084696
```

```
##  [1] "13-14" "13-15" "13-16" "13-17" "14-15" "14-16" "14-17" "15-16"
##  [9] "15-17" "16-17"
```

```
(p.yr <- data.frame(yrs, D, p.val))
```

```
##      yrs          D       p.val
## 1  13-14 0.23721016 0.014291577
## 2  13-15 0.19517544 0.344309858
## 3  13-16 0.09997322 1.000000000
## 4  13-17 0.14160401 1.000000000
## 5  14-15 0.19003497 0.344309858
## 6  14-16 0.24747771 0.004613498
## 7  14-17 0.20313020 0.686011563
## 8  15-16 0.24265267 0.046126129
## 9  15-17 0.13630952 1.000000000
## 10 16-17 0.17084696 1.000000000
```

Summary of Results

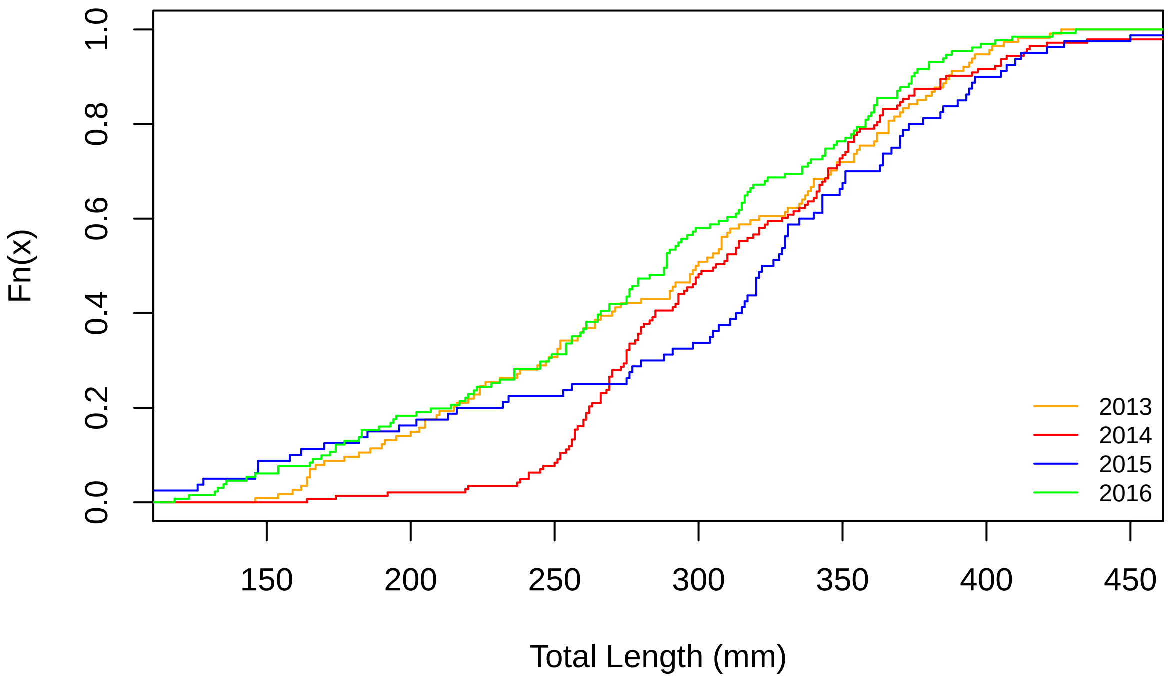
The results of the Kolmogorov-Smirnov test above seem to suggest the largemouth bass population is not stable (*Or is it? I think there are just a few weird years probably sampling related*). The length frequency distribution **is significant different** between the years 2013 and 2014 ($D = 0.24$, $P = 0.014$), 2014 and 2016 ($D = 0.25$, $P < 0.005$), and 2015 and 2016 ($D = 0.24$, $P = 0.046$). There is **no significant differnce** between the length frequency distributions for 2013 and 2015 ($D = 0.20$, $P = 0.344$), 2013 and 2016 ($D = 0.10$, $P = 1$), and 2014 and 2015 ($D = 0.19$, $P = 0.344$). The length frequency distribution for the year 2017 was *not significantly different* between any years.

Note: *Adding in the Year 2017 Significantly Altered the Adjusted P-Values*

Length Frequency 2013 - 2016 (without 2017)

Cumulative Frequencies 2013 - 2016

Lets look at the empirical cumulative distribution function (ECDF). This is the proportion of fish less than each observed length. This should help me compare the length frequency distributions between years.



Compare Length Frequency Between Years

Kolmogorov-Smirnov Test

```
## [1] 0.23721016 0.19517544 0.09997322 0.19003497 0.24747771 0.24265267
```

```
## [1] "13-14" "13-15" "13-16" "14-15" "14-16" "15-16"
```

```
(p.yr <- data.frame(yrs, D, p.val))
```

```
##     yrs          D       p.val
## 1 13-14 0.23721016 0.007939765
## 2 13-15 0.19517544 0.147561368
## 3 13-16 0.09997322 0.576122817
## 4 14-15 0.19003497 0.147561368
## 5 14-16 0.24747771 0.002768099
## 6 15-16 0.24265267 0.023063064
```

Summary of Results

The results of the Kolmogorov-Smirnov test above seem to suggest the largemouth bass population is not stable (*Or is it? I think there are just a few weird years probably sampling related*). The length frequency distribution **is significant different** between the years 2013 and 2014 (D = 0.24, P = 0.008),2014 and 2016 (D = 0.25, P = 0.003), and 2015 and 2016 (D = 0.24, P = 0.023). There is **no significant difference** between the length frequency distributions for 2013 and 2015 (D = 0.20, P = 0.148), 2013 and 2016 (D = 0.10, P = 0.576), and 2014 and 2015 (D = 0.19, P = 0.148).

Length Frequency 2013 - 2016 (No Fish < 200mm)

What is I remove small fish (<200mm) which I am not sure are fully sampled by my gear. Will 2014 still be different from 2013 and 2016? will 2015?

2014 has a lot of stock length individuals (n stock = 65) compared to 2013 (41), 2015 (14), and 2016 (52) but far fewer substock individuals (3), 2013 (16), 2015 (13), 2016 (24). 2014 also seems to have a bit more but comperable amount of quality and preferred length individuals (57,18 respectivley) than 2013 (41,16), 2015 (38,15), and 2016 (44,11). So lets explore wether the differences are in the smaller sub stock length not fully sampled by our gear.

```
lmb.s <- read.csv("Data/Clean-Data/2012-2017_nearshore-survey-largemouth-bass_Stock_CLEAN.csv") %>%
    arrange(Year, FID, Length)
lmb.s$fyr <- as.factor(lmb.s$fyr)
```

```
unique(lmb.s$Year) ### See that there is no 2013
```

```
## [1] 2012 2013 2014 2015 2016 2017
```

```
table(lmb.s$gcat) ### Check that there are no sub-stock fish
```

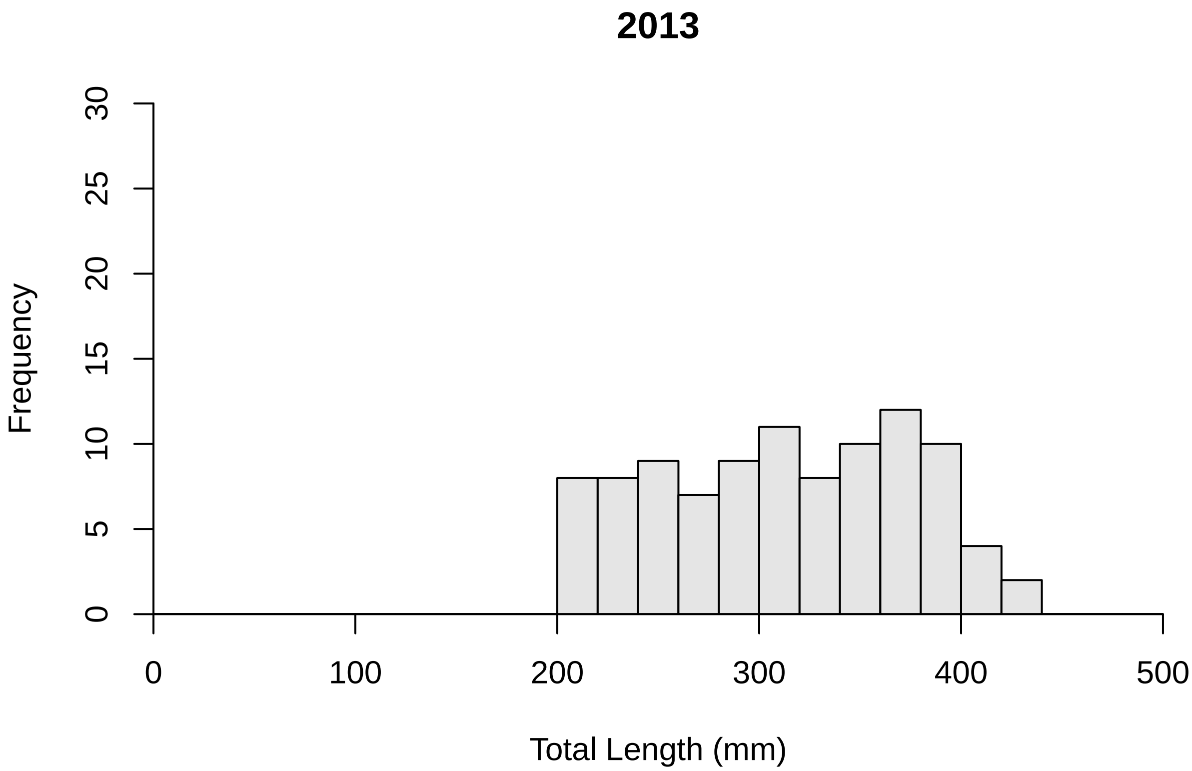
```
##
## preferred   quality     stock
##        79       203       186
```

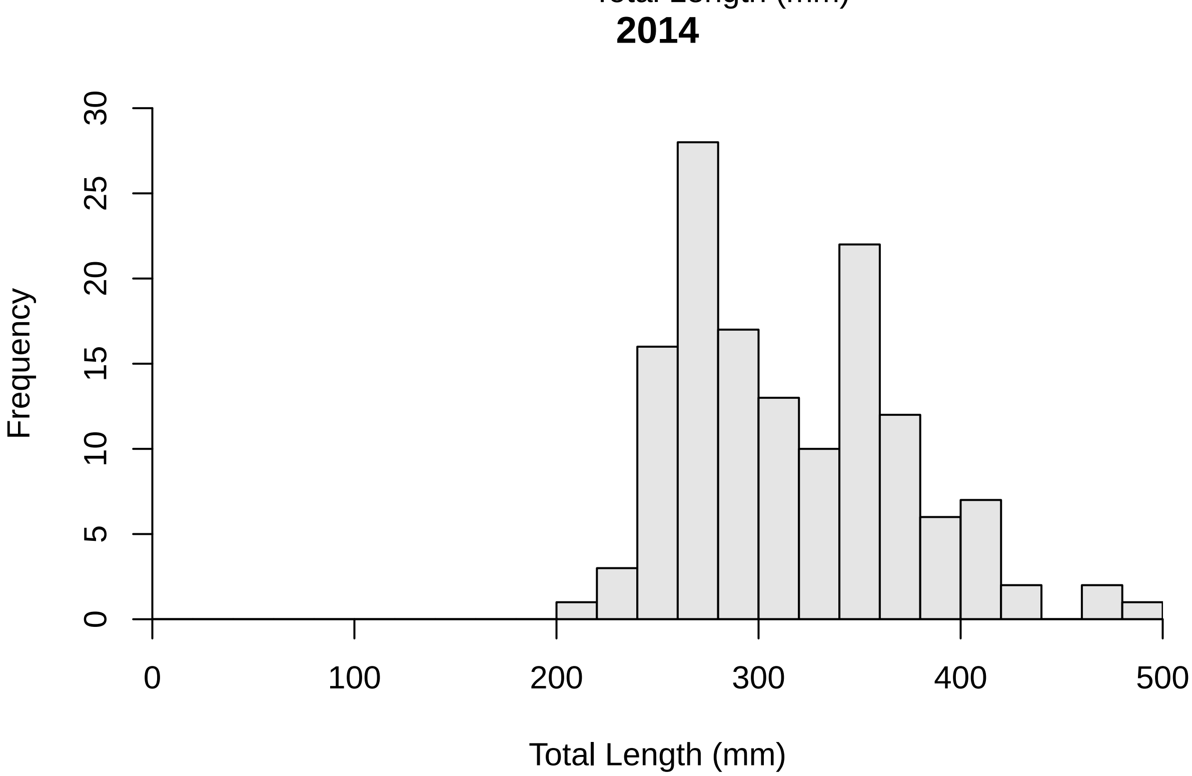
```
lmb.s %<>% mutate(lcat20 = lencat(Length, w = 20))
```

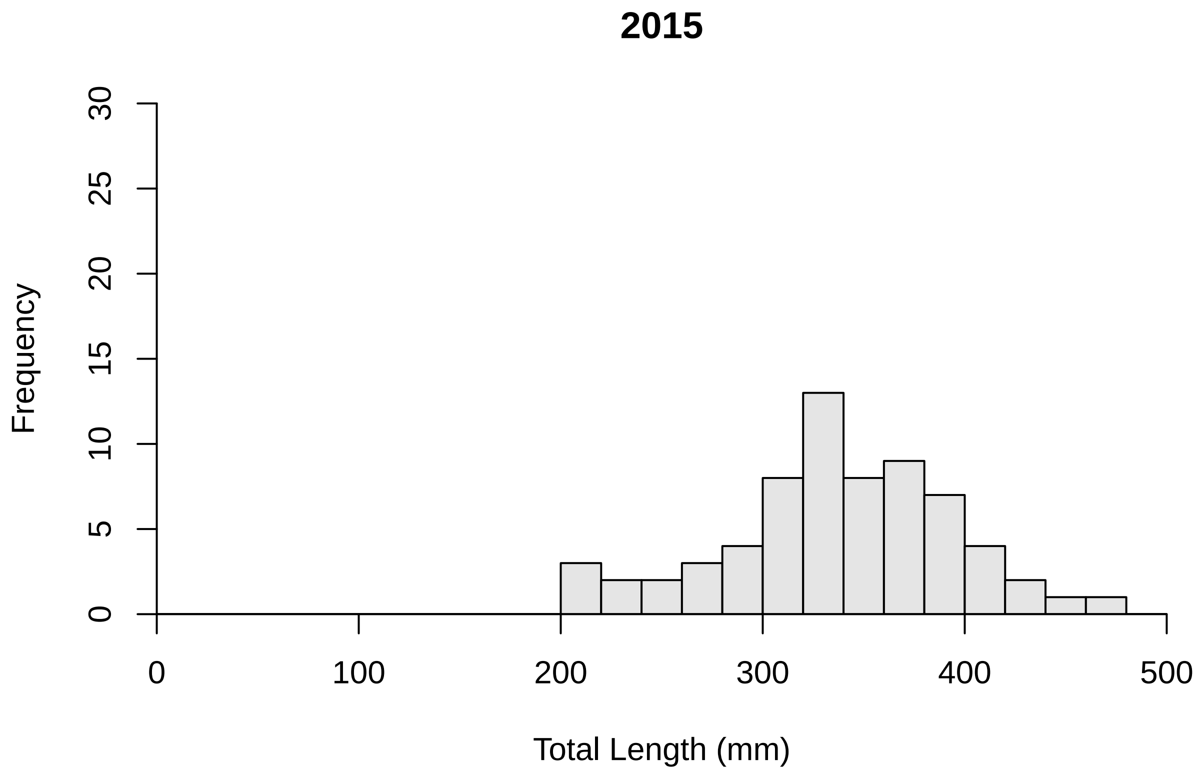
```
lmbs.12 <- filter(lmb.s, Year == 2012)
lmbs.13 <- filter(lmb.s, Year == 2013)
lmbs.14 <- filter(lmb.s, Year == 2014)
lmbs.15 <- filter(lmb.s, Year == 2015)
lmbs.16 <- filter(lmb.s, Year == 2016)
lmbs.17 <- filter(lmb.s, Year == 2017)
```

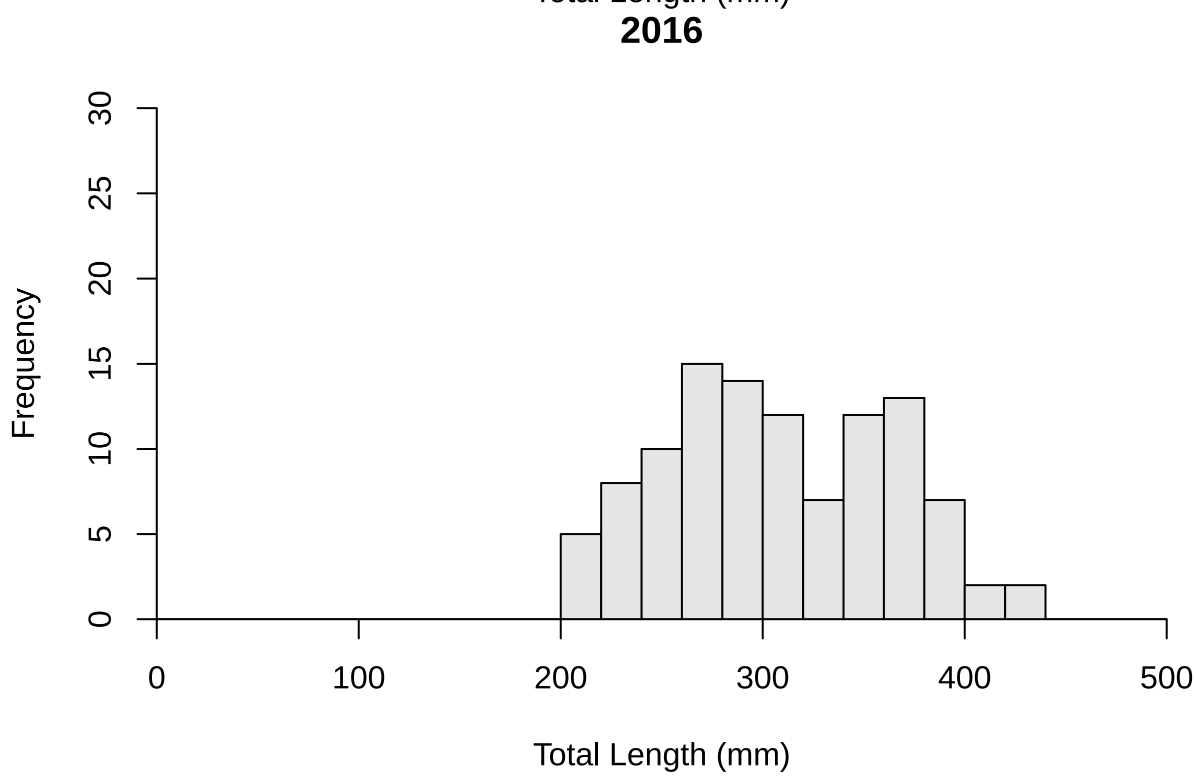
Length Frequency Distribution

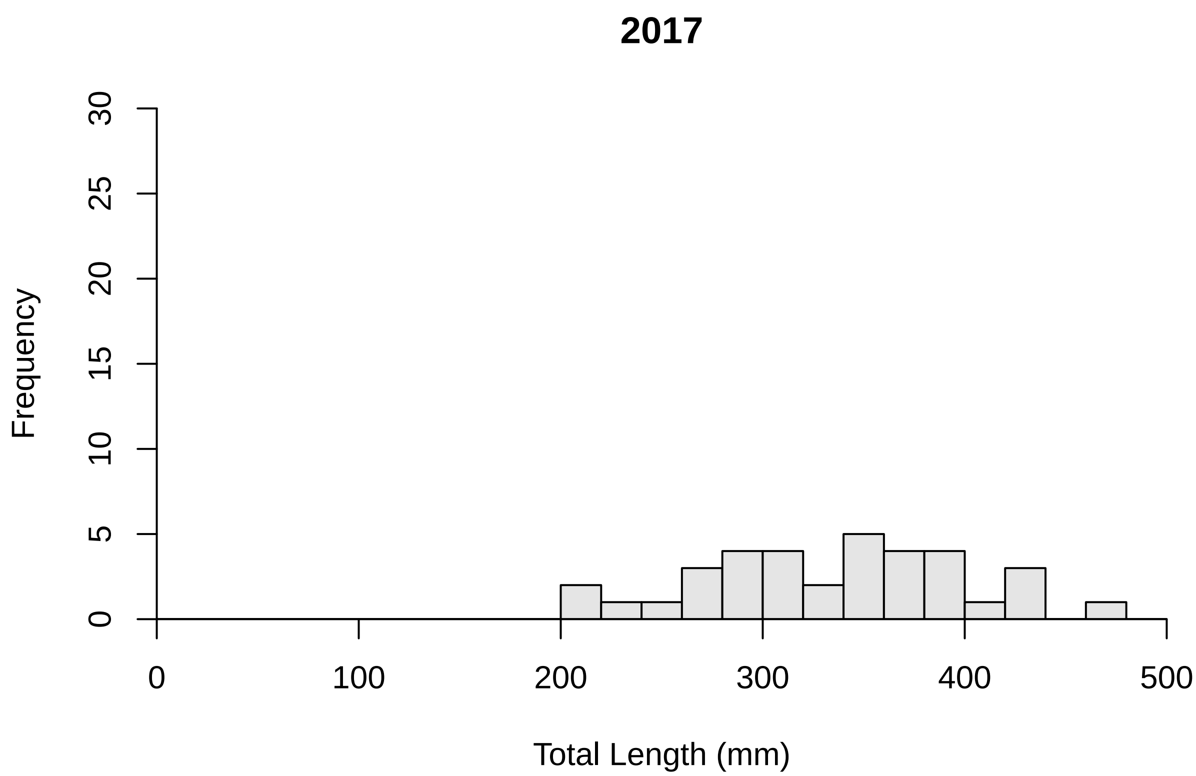
Lets view a quick histogram of the frequency of fish in each length bin.


2013
Frequency
0
5
10
15
20
25
30
0
100
200
300
400
500
Total Length (mm)


2014
Frequency
0
5
10
15
20
25
30
0
100
200
300
400
500
Total Length (mm)

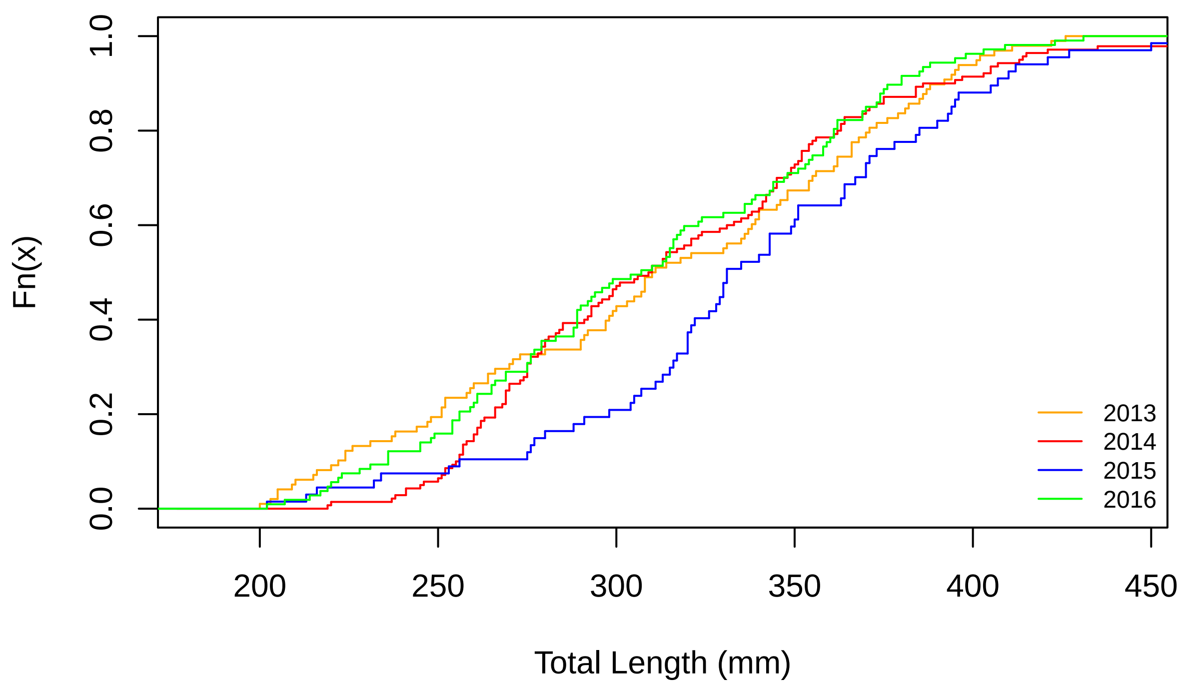

2015
Frequency
0
5
10
15
20
25
30
0
100
200
300
400
500
Total Length (mm)


2016
Frequency
0
5
10
15
20
25
30
0
100
200
300
400
500
Total Length (mm)



Cumulative Frequencies (Stock)

Lets look at the empirical cumulative distribution function (ECDF). This is the proportion of fish less than each observed length. This should help me compare the length frequency distributions between years.



Compare Length Frequency Between Years (for stock only)

Kolmogorov-Smirnov Test (Stock)

```
## [1] 0.1489796 0.2462687 0.0897387 0.2696162 0.1072096 0.2770261
```

```
## [1] "13-14" "13-15" "13-16" "14-15" "14-16" "15-16"
```

```
(p.yr.s <- data.frame(yrs, D.s, p.val.s))
```

```
##     yrs       D.s    p.val.s
## 1 13-14 0.1489796 0.46410352
## 2 13-15 0.2462687 0.06409271
## 3 13-16 0.0897387 0.97704002
## 4 14-15 0.2696162 0.01652127
## 5 14-16 0.1072096 0.97704002
## 6 15-16 0.2770261 0.01793051
```

Summary of Results

The results of the Kolmogorov-Smirnov test above seem to suggest the largemouth bass population is not stable (*Or is it? I think there are just a few weird years probably sampling related*). The length frequency distribution **is significant different** between the years 2014 and 2015 (D = 0.27, p = 0.017), and 2015 and 2016 (D = 0.28, p = 0.018). There is **no significant difference** between the length frequency distributions for 2013 and 2014 (D = 0.15, p = 0.464), 2013 and 2015 (D = 0.25, P = 0.064), 2013 and 2016 (D = 0.09, P = 0.977), and 2014 and 2016 (D = 0.11, P = 0.977).